

Please type a plus sign (+) inside this box ☐

Approved for use through 09/30/00 OMB 0651-0032
Patent and Trademark Office U S DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number

**UTILITY
PATENT APPLICATION
TRANSMITTAL**

Attorney Docket No.: CITI0168

Total Pages:

First Named Inventor or Application Identifier
Ronald COLEMAN, et al.

Express Mail Label No.:

APPLICATION ELEMENTS

See MPEP chapter 600 concerning utility patent application contents

ADDRESS TO: Commissioner for Patents
Box Patent Application
Washington, D.C. 20231

1. ☒ Fee Transmittal Form
(Submit an original, and a duplicate for fee processing)
2. ☒ Specification Total Pages **24**
(preferred arrangement as set forth below)
 - Descriptive title of the invention
 - Cross References to Related Applications
 - Statement Regarding Fed sponsored R & D
 - Reference to Microfiche Appendix
 - Background of the invention
 - Brief Summary of the invention
 - Brief Description of the Drawings (if filed)
 - Detailed Description
 - Claim(s)
 - Abstract of the Disclosure
3. ☒ Drawing(s) (35 usc 113) [Total Sheets] **10**
4. ☐ Oath or Declaration [Total Pages] ☐
 - a. ☐ Newly executed (original or copy)
 - b. ☐ Copy from a prior application (37 CFR 1.63(d)) (for continuation/divisional with box 17 completed) [Note Box 5 below]
 - i. ☐ **DELETION OF INVENTOR(S)**
Signed statement attached deleting inventor(s) named in the prior application. See 37 CFR 1.63(d)(2) and 1.33(b).
5. ☐ Incorporation By Reference (usable if Box 4b is checked)
The entire disclosure of the prior application, from which a copy of the oath or declaration is supplied under Box 4b, is considered as being part of the disclosure of the accompanying application and is hereby incorporated by reference therein.
6. ☐ Microfiche Computer Program (Appendix)
7. Nucleotide and/or Amino Acid Sequence Submission (if applicable, all necessary)
 - a. ☐ Computer Readable Copy
 - b. ☐ Paper Copy (identical to computer copy)
 - c. ☐ Statement verifying identify of above copies

ACCOMPANYING APPLICATION PARTS

8. ☐ Assignment Papers (cover sheet & document(s))
9. ☐ 37 CFR 3.73(b) Statement ☐ Power of Attorney (when there is an assignee)
10. ☐ English Translation Document (if applicable)
11. ☐ Information Disclosure Statement (IDS)/PTO-1449 ☐ Copies of IDS Citations
12. ☐ Preliminary Amendment
13. ☒ Return Receipt Postcard (MPEP 503)
14. ☐ Small Entity Statement ☐ Statement filed in prior application, Status till proper and desired
15. ☐ Certified Copy of Priority Document(s) (If foreign priority is claimed)
16. ☐ Other: _____

17. If a **CONTINUING APPLICATION**, check appropriate box and supply the requisite information:

☐ Continuation ☐ Divisional ☐ Continuation-in-part (CIP) of prior application No.: _____

18. CORRESPONDENCE ADDRESS

☐ Customer Number or Bar Code Label or ☐ Correspondence address below

Name	George T. Marcou				
	KILPATRICK STOCKTON LLP				
Address	700 13th Street, N.W.				
	Suite 800				
City	Washington	State	DC	Zip Code	20005
Country	U.S.A.	Telephone	202.508.5800	Fax	202.508.5858
Name (Print/Type)	George T. Marcou			Reg. No.	33,041
Signature	T. Marcou			Date	8/8/00

Burden Hour Statement. This form is estimated to take 0.2 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, Patent and Trademark Office, Washington, D.C. 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Box Patent Application, Washington, D.C. 20231.

SYSTEM AND METHOD FOR ASSURING THE INTEGRITY OF DATA USED TO EVALUATE FINANCIAL RISK OR EXPOSURE

This application claims priority to co-pending provisional application entitled
"CONTENT ANALYSIS" having U.S. Serial Number 60/147,487 filed August 9, 2000.

Field of the Invention

The present invention relates to a system and method for measuring the financial risks associated with trading portfolios. Moreover, the present invention relates to a system and method for assuring the integrity and validity of data used to evaluate financial risk or exposure.

Background of the Invention

As companies and financial institutions grow more dependent on the global economy, the volatility of currency exchange rates, interest rates, and market fluctuations creates significant risks. Failure to properly quantify and manage risk can result in disasters such as the failure of Barings ING. To help manage risks, companies can trade derivative instruments to selectively transfer risk to other parties in exchange for sufficient consideration.

A derivative is a security that derives its value from another underlying security. Derivatives also serve as risk-shifting devices. Initially, they were used to reduce exposure to changes in independent factors such as foreign exchange rates and interest rates. More recently, derivatives have been used to segregate categories of investment risk that may

appeal to different investment strategies used by mutual fund managers, corporate treasurers or pension fund administrators. These investment managers may decide that it is more beneficial to assume a specific risk characteristic of a security.

Derivative markets play an increasingly important role in contemporary financial markets, primarily through risk management. Derivative securities provide a mechanism through which investors, corporations, and countries can effectively hedge themselves against financial risks. Hedging financial risks is similar to purchasing insurance; hedging provides insurance against the adverse effect of variables over which businesses or countries have no control.

Many times, entities such as corporations enter into transactions that are based on a floating rate, interest, or currency. In order to hedge the volatility of these securities, the entity will enter into another deal with a financial institution that will take the risk from them, at a cost, by providing a fixed rate. Both the interest rate and foreign exchange rate derivatives lock in a fixed rate/price for the particular transaction one holds.

For example, Alan loans Bob \$100 dollars on a floating interest rate. The rate is currently at 7%. Bob calls his bank and says, "I am afraid that interest rates will rise. Let us say I pay you 7% and you pay my loan to Alan at the current floating rate." If rates go down, the bank makes the money on the spread (the difference between the 7% float rate and the new lower rate) and Bob is borrowing at a higher rate. If rates rise however, then the bank loses money and Bob is borrowing at a lower rate. Banks usually charge a risk/service fee, in addition, to compensate for the additional risk.

Consider another example: If ABC, an American company, expects payment for a shipment of goods in British Pound Sterling, it may enter into a derivative contract with Bank A to reduce the risk that the exchange rate with the U.S. Dollar will be more unfavorable at the time the bill is due and paid. Under the derivative instrument, Bank A is obligated to pay ABC the amount due at the exchange rate in effect when the derivative contract was executed. By using a derivative product, ABC has shifted the risk of exchange rate movement to Bank A.

The financial markets increasingly have become subject to greater “swings” in interest rate movements than in past decades. As a result, financial derivatives have also appealed to corporate treasurers who wish to take advantage of favorable interest rates in the management of corporate debt without the expense of issuing new debt securities. For example, if a corporation has issued long term debt with an interest rate of 7 percent and current interest rates are 5 percent, the corporate treasurer may choose to exchange (i.e., swap) interest rate payments on the long term debt for a floating interest rate, without disturbing the underlying principal amount of the debt itself.

In order to manage risk, financial institutions have implemented quantitative applications to measure the financial risks of trades. Calculating the risks associated with complex derivative contracts can be very difficult, requiring estimates of interest rates, exchange rates, and market prices at the maturity date, which may be twenty to thirty years in the future. To make estimates of risk, various statistical and probabilistic techniques are used. These systems, called Pre-Settlement Exposure Servers (PSE Servers) are commonly known in the art.

PSE Servers simulate market conditions over the life of the derivative contracts to determine the exposure profile representing the worst case scenario within a 97.7% confidence interval, or approximately two standard deviations. This exposure profile is calculated to give current estimates of future liabilities. As market conditions fluctuate from day to day or intra-day, the calculated exposure profile changes; however, these changes are not always due to market fluctuations, they are sometimes due to errors in the input data.

In the past, input data errors have been manually detected by users; however, since the quantity of input data is now so large, it is impossible for users to detect and correct all of the errors. Users are most likely to detect errors in the input data that cause a significant change in the exposure profile.

The present invention seeks to automatically detect errors in input data to the PSE Server using an information theory technique known as Content Analysis. Content Analysis, based on information theory, attempts to look for sweeping changes or statistically significant trends in data suggestive of error. If statistically significant changes are detected, users can be alerted that one or more errors in the input data is likely. This prevents invalid data from skewing the resulting exposure profiles, providing more accurate estimations of possible exposure.

Summary of the Invention

In accordance with the invention, a method and system are provided for detecting abnormalities in input data to a financial risk management system. The method includes receiving a set of input data to a financial risk management system; receiving one or more

historical values, each historical value representing a calculated content from a previous set of input data; and calculating the likelihood that changes to the set of input data are the result of one or more errors.

In further aspects of the invention, the input data includes data feeds from one or more data processing system as well as calculated data from a financial risk management system. In one embodiment of the invention, a result is determined based on the calculated likelihood that changes to the set of input data are the result of one or more errors. The result is then displayed. In one embodiment of the present invention, the result is displayed to users as an icon indicative of the degree of likelihood that changes to the set of input data are the result of one or more errors.

In yet a further aspect of invention, the likelihood that changes to the set of input data are the result of one or more errors is calculated by determining the information content of the input data, and performing a statistical analysis of the calculated information content relative to historical values to determine the likelihood that changes to the input data are the result of one or more errors. The information content of input data can be calculated by determining the Shannon entropy of the data and the statistical analysis can be performed using non-parametric statistics, parametric statistics, or Bayesian statistics.

Brief Description of the Drawings

Having thus briefly described the invention, the same will become better understood from the following detailed discussion, taken in conjunction with the drawings where:

FIG. 1 is a network diagram showing a PSE Server according to one embodiment of the present invention;

FIG. 2 is pseudocode describing the calculation of Ω for discrete data inputs according to an embodiment of the present invention;

FIG. 3 is pseudocode describing the calculation of Ω for continuous data inputs according to one embodiment of the present invention;

FIG. 4 is pseudocode describing the calculation of Ω for continuous by continuous data inputs according to one embodiment of the present invention;

FIG. 5 is pseudocode describing the calculation of Ω for continuous by discrete data inputs according to one embodiment of the present invention;

FIG. 6 is pseudocode describing the calculation of Ω for discrete by discrete data inputs according to one embodiment of the present invention;

FIG. 7 is a table depicting semaphores representing the likelihood of errors according to an embodiment of the present invention;

FIG. 8 is a screenshot depicting the results of applying Content Analysis to input data according to an embodiment of the present invention;

FIG. 9 is a diagram describing the handling of boundary conditions while performing Content Analysis on continuous input data according to one embodiment of the present invention; and

FIG. 10 is a flow chart describing a method for identifying input errors in input data according to an embodiment of the present invention.

Detailed Description

In the late 1940s, Claude Shannon, an American engineer working for Bell Telephone Labs, made a monumental discovery---the connection between physical entropy and *information entropy*. Shannon understood that the amount of "information" in a message is its entropy. Entropy is exactly the amount of information measured in *bits* needed to send a message over the telephone wire or, for that matter, any other channel including the depths of space. At maximum entropy, a message is totally incomprehensible, being random gibberish, containing no useful information.

The present invention uses a method we call Content Analysis to determine if changes in financial information are likely the result of errors. Content Analysis uses the Shannon measure of information content; however, instead of working with messages, Content Analysis works with financial information. Much financial information is far from equilibrium, meaning the data is highly non-normally distributed. Thus this condition, while not readily suitable for ordinary statistics, is ideal for entropy analysis. We call our measurement of content not entropy but *omega* (Ω).

Content Analysis consists of two parts: (1) first, trading information is *thermalized* by converting it to Shannon entropy; and (2) then, the resulting data is processed further by applying statistical analysis to determine if changes are likely caused by errors in input data. In the preferred embodiment of the present invention, the thermalized data is processed using non-parametric resampling statistics on changes in content. Given a change in content, non-

parametric resampling statistics provide a mechanism to deduce the probability of a Type I Error at a given statistical confidence level.

Additional embodiments of the present invention use other statistical methods commonly known in the art. Any method that can determine whether the thermalized change is likely the result of one or more errors instead of expected fluctuations in market conditions or changed positions can be used to perform Content Analysis. For example, alternative statistics such as parametric or Bayesian statistics can be used. The preferred embodiment of the present invention uses resampling statistics because they are robust and they are easy to use and implement. The only potential drawback to resampling statistics is speed; though in practice modern computer processors are fast enough to provide adequate performance.

Content Analysis determines the confidence level that a change in input trading data is caused by errors. This confidence level is then presented on a logarithmic scale of odds ratios which we call the maximum credible assessment. Our assessment scale is attributed to Harold Jefferys, a British geophysicist and pioneering statistician of the Bayesian school of the 1930s.

There are several applications and benefits to looking at trading information in this way. One advantage is that the description of complex financial data, both trading contracts and spot market factors, is standardized in terms of actual content. Thus, different quantities can be compared and discussed meaningfully using a more abstract but measurable quantity, although representing disparate information. Once in standard form, statistics, numerical analysis, etc. can be run against the data.

Thus, we are mainly interested in $\Delta\Omega$ (i.e., changes in information content). The difference is analogous to measuring the temperature of a heat bath versus measuring changes in temperature of the heat bath. Given $\Delta\Omega$, we can compile historical data and look for unexpected fluctuations as a plausible indication that the data integrity has been compromised. Now that Content Analysis has been described generally, we now turn to a detailed description of an implementation according to a preferred embodiment of the present invention.

FIG. 1 is a network diagram showing a PSE Server 101 attached to a computer network 102. The PSE Server 101 uses techniques commonly known in the art to determine an exposure profile representing the worst case scenario within a two standard deviation confidence interval (i.e., 97.7% confidence). In the preferred embodiment, the data calculations made by the PSE Server 101 are stored on the computer system as a file that can be accessed by a software application according to the present invention.

The PSE Server 101 collects data from various sources regarding portfolios of derivative instruments. Using the collected data, the PSE Server 101 derives and or receives various measurements of exposure or risk such as the Current Mark to Market ("CMTM") and the Maximum Likely Increase in Value ("MLIV"). The CMTM is the current market value of a portfolio of financial instruments and the MLIV is the maximum likely increase in value of a trade.

One embodiment of the present invention uses a data file containing the results from conventional calculations performed by the PSE Server 101 to perform Content Analysis and

thus determine whether changes in the exposure profile are likely caused by some error in the input data. Before describing how the present invention uses Content Analysis, we must first describe how the content of various kinds of information is calculated.

Table 1 gives the mathematical formulae for calculating Ω for each object type. An object is just a measurable quantity of information in the Server. For example, product codes, zero coupon discount curves, etc. The total number of objects in the macrostate (the universe of objects) is always N and each microstate (a sub-universe) has N_i objects. Objects may be discrete (e.g., product codes) or continuous (e.g., CMTMs). The number of microstates for discrete objects is M or M_1 and M_2 . The number of microstates for continuous objects is a function of the number of dimensions and the object type(s). We choose N_i in such a way so that the search complexity is reasonable. This number N_i is justified by an empirical analysis of the current size of the global book for the largest counterparty and the expected growth over the foreseeable future.

Thus, for the continuous case, we choose $N_i = \lceil \sqrt{N} \rceil$. For the continuous \times continuous case, we choose $N_i = \lceil \sqrt[4]{N} \rceil$. For the continuous \times discrete case, we have $\alpha = \log M / \log N$ so that $N_i = \lceil N^\alpha \rceil$ where $0 < \alpha \leq 1$. In the continuous cases, boundary conditions are handled. This is shown for one dimension in Figure 9.

Table 1

Type(s)	Ω	Ω_{\max}	N_{\min}
discrete	$\sum_i^M N_i \log N / N_i$	$N \log M$	2

discrete \times discrete	$\sum_i^{M_1} \sum_j^{M_2} N_{i,j} \log N / N_{i,j}$	$N \log M_1 M_2$	4
continuous	$\sum_i^{\lceil \sqrt{N} \rceil} N_i \log N / N_i$	$N \log \sqrt{N}$	4
continuous \times continuous	$\sum_i^{\lceil \sqrt{N} \rceil} \sum_j^{\lceil \sqrt{N} \rceil} N_{i,j} \log N / N_{i,j}$	$N \log \sqrt{N}$	16
continuous \times discrete	$\sum_i^{N^\alpha} \sum_j^M N_{i,j} \log N / N_{i,j}$	$N \log N^\alpha M$	$N^\alpha M = 2$

Table 1 describes how content analysis is performed using five modes of input data: discrete, discrete \times discrete, continuous, continuous \times continuous, and continuous \times discrete. FIGS. 2-6 describe a method for computing Ω for each mode of input data using pseudocode. One skilled in the art will appreciate that each of these methods described by FIGS. 2-6 can be easily implemented in most modern computer languages. In the preferred embodiment of the present invention, a Perl script is used to read the input data from the PSE Server 101 and to perform Content Analysis.

Using these techniques to compute the information content of the input data, the following reports described below in Table 2 can be generated with the data from the PSE Server: (1) CMTM; (2) CMTM \times Product; (3) MLIV; (4) MLIV \times Product; (5) Fails; (6) Fails \times Product; (7) Bad; (8) Bad \times Product; (9) Netting; (10) Products; (11) Netting Product; (12) CMTM \times MLIV; (13) Passes; and (14) Passes \times Product, where CMTM is the "Current Mark to Market" and MLIV is the "Most Likely Increase in Value". In one

embodiment of the present invention, these fourteen Content Analysis reports are displayed in a grid as shown in FIG. 8. The report grid is designed to provide a comprehensive picture of how content across counter-parties is changing. Thus, if there is a detectable trend, it should be fairly easy to spot the pattern.

Table 2

Feature Content	Comment
CMTM	This analysis measures changes in CMTM over all trades for the counter-party. The analysis holds potential to reveal content shifts in the portfolio as a hold.
CMTM by Product	This analysis measures changes in CMTM over all trades by product for the counter-party. The analysis holds potential to reveal content shifts that are isolated to a product group.
MLIV	This analysis measures changes in MLIV over all trades, pass or fail, for the counter-party. The analysis holds potential to reveal content shifts in the portfolio.
MLIV by Product	This analysis measures changes in MLIV over all trades by product for the counter-party. The analysis holds potential to reveal content shifts that are isolated to a product group.
CMTM by MLIV	This analysis measures changes in CMTM over all trades by MLIV for the counter-party. It may perhaps be a little difficult to visualize this in two dimension but imagine a scatter plot of CMTM and MLIV. The analysis holds potential to reveal content shifts that are isolated to one or more areas of the scatter.
Netting	This analysis measures changes in the netting structure over all trades for the counter-party. The analysis holds potential to reveal content shifts in the netting of a portfolio that is not detectable by just looking at the total netting count.
Netting by Product	This analysis measures changes in the netting structure over all trades by netting agreement for the counter-party. The visualization problem here is the same as CMTM and MLIV: namely, try to imagine a scatter plot of netting agreements and products. The analysis hold potential to reveal content shifts that are isolated to one or more areas of the scatter.
Product	This analysis measures changes in products over all trades for the counter-party. The analysis holds potential to reveal content shifts in the portfolio of products.
Passed	This analysis measures changes in pass counts over all trades for the counter-party. The analysis holds potential to reveal pass count shifts over all trades in the portfolio.

Passed by Product	This analysis is very similar the analysis for products; here the content is filtered only for products that pass the tolerance test.
Failed	This analysis measures changes in fail counts over all trades for the counter-party. The analysis holds potential to reveal fail count shifts over all trades in the portfolio.
Failed by Product	This analysis is very similar the analysis for products; here the content is filtered only for products that fail the tolerance test. The analysis holds potential to reveal content shifts isolated to failed products.
Bad	This analysis measures changes in bad counts over all trades for the counter-party. The analysis holds potential to reveal bad count shifts over all trades in the portfolio.
Bad by Product	This analysis is very similar the analysis for products; here the content is filtered to capture bad products. The analysis holds potential to reveal contents shifts isolated to bad products.

The following table describes some of the reports that can be generated using Content Analysis as well as whether the feature measured is continuous, discrete, or a combination of the two. These reports are displayed in a graphical user interface such as that shown in FIG. 8. using the semaphores. A user can use the report displayed by the graphical user interface to determine if there are errors in the data that need attention.

Table 3

Feature	Discrete or Continuous	Basic or Complex
Net agreements	Discrete	Basic
Products	Discrete	Basic
Schedule records	Discrete	Basic
Time to maturity	Continuous	Basic
CMTMs	Continuous	Basic
MLIVs	Continuous	Basic
Net agreements × Products	Discrete-Discrete	Complex

Net agreements × CMTMs	Discrete-Continuous	Complex
CMTM × MLIV	Continuous-Continuous	Complex

The present invention uses these reports to determine where human intervention is likely to be necessary. Thus, users can be alerted to the possibility of bad data and shown the input data that has substantially different information content than historical runs. This information can be displayed in a graphical user interface using the symbols shown in FIG. 7.

The goal of Content Analysis is to put changes in content, not content per se, into perspective. The idea of Content Analysis rests on a premise so obvious it is often overlooked: namely, that data feeds are in a constant state of flux. The problem, however, is that sometimes manual inspection fails to distinguish between “normal” changes we might expect from ordinary business/systems operations versus data errors caused by those operations, including human faults, system failures, and whatnot.

Content Analysis assesses changes in content using a simple odds scale called *maximum credible assessments*. The maximum credible assessment gives the most we could say in practice about content changes which we categorize as normal, outer normal, borderline, and abnormal changes. The maximum credible assessment criteria are summarized in Table 4 below. These criteria are arbitrary; one of ordinary skill in the art will appreciate that these values can be modified without departing from the spirit of the present invention. Additional embodiments of the present invention can include varying

numbers of change categories. For example, a three category system can be provided including the following change categories: Normal, Borderline, and Abnormal.

Table 4

Change	Odds favoring problem	Potential of problem (Maximum credible assessment)
Normal	3 to 1	Little potential of problem
Outer Normal	6 to 1	Substantial potential of problem
Borderline	20 to 1	Strong potential of problem
Abnormal	>20 to 1	Decisive potential of problem

As shown in Table 4, changes to trading data is likely. Since some change is expected and not necessarily the result of errors, we select ranges of odds that are indicative of errors to the input data. In other applications, input data may be more regular than in the present embodiment. If data is more regular, then smaller changes in content may be more likely caused by errors than that shown in Table 4.

In other words, the maximum credible assessment is only a statement of plausibility, not actuality. The maximum credible assessments have been designed so that we really only have to worry about two kinds of changes: borderline and abnormal. These represent "big" or "near-big" changes in content.

Content Analysis measures changes in content relative to expectations based on recent history. This is a loaded statement, the importance of which cannot be emphasized enough. Essentially the change categories listed in Table 4 are not static, predefined ideals. They are measurements relative to our expectations based on historic or prior data which are always changing as feeds change. The likelihood that a change is abnormal is a measure of the change relative to the prior history of data feed. Content Analysis is not only measuring changes in the content or Ω of input data, but it also measures the likelihood that the changes are abnormal. Thus, the statistics of Content Analysis are regularly changing based on historic data feeds. Consequently what is a normal change in content today might not be normal next week depending on recent history.

Recent history is essentially a sliding window of feeds which we use to compute the statistics of Content Analysis as far as expectations go. The size of the sliding window itself is two to three weeks depending on a couple of factors.

Factor one concerns how feeds have come into the Server. If feeds have been missed, i.e., not sent to the Server, the sliding window of recent history shrinks one day. If feeds are not sent for two days in a row, recent history shrinks by two days and so on.

Factor two concerns how feeds have been released. If an entire feed is canceled, we have the same situation as Factor One. If, however, a counter-party is canceled, we have a different situation in which the window remains the same size but the content is slightly skewed for the counter-party. This occurs because performing release-by-counter-party makes the system use the last known data believed to be good for the current run. Inside the

Server this means the feed for the counter-party is duplicated (or triplicated if a counter-party is canceled twice in a row) which tends to distort the content.

Distorted content caused by a shrinking window of historical data or by duplicated or triplicated data, tends to make Content Analysis more sensitive to content changes. A change that would have been normal otherwise, may move in the outer normal direction as repeated historical data amplifies any changes that may occur.

Fortunately, resampling statistics are robust enough to gracefully handle these problems. Moreover, the window distortions eventually correct themselves as old feeds are removed from the system. The sliding window reverts to its normal size and content distortions are minimized.

Embodiments of the present invention have now been generally described in a non-limiting manner. It will be appreciated that these examples are merely illustrative of the present invention, which is defined by the following claims. Many variations and modifications will be apparent to those of ordinary skill in the art.

Claims

What is claimed is:

1. A method for detecting abnormalities in input data to a financial risk management system, the method comprising:

- (a) receiving a set of input data to a financial risk management system;
- (b) receiving one or more historical values, each historical value representing a previous set of input data;
- (c) calculating the likelihood that changes to the set of input data are the result of one or more errors.

2. The method of claim 1, wherein the input data includes data feeds from one or more data processing systems.

3. The method of claim 1, wherein the input data includes data calculated by a financial risk management system.

4. The method of claim 1, further comprising:

- (d) displaying a result based on the calculated likelihood that changes to the set of input data are the result of one or more errors.

5. The method of claim 4, wherein displaying a result includes displaying an icon indicative of the degree of likelihood that changes to the set of input data are the result of one or more errors.

6. The method of claim 1, wherein calculating the likelihood that changes to the set of input data are the result of one or more errors comprises:

- (i) calculating the information content of the input data; and
- (ii) performing a statistical analysis of the calculated information content relative to the one or more historical values to determine the likelihood that changes to the input data are the result of one or more errors.

7. The method of claim 6, wherein calculating the information content of the input data is performed by calculating the Shannon entropy of the input data.

8. The method of claim 6, wherein the statistical analysis is performed using non-parametric resampling statistics.

9. The method of claim 6, wherein the statistical analysis is performed using Bayesian statistics.

10. The method of claim 6, wherein the statistical analysis is performed using parametric statistics.

11. A system for detecting abnormalities in input data to a financial risk management system, the system comprising:

a data processing server that receives a set of input data;

a computer storage device for storing one or more historical values, each historical value representing a previous set of input data; and

one or more central processing units coupled to the computer storage device, the one or more central processing units calculating the likelihood that changes to the set of input data are the result of one or more errors.

12. The system of claim 11, wherein the input data includes data feeds from one or more data processing systems.

13. The system of claim 11, wherein the input data includes data calculated by a financial risk management system.

14. The system of claim 11, further comprising:

a graphical user interface that displays a result based on the calculated likelihood that changes to the set of input data are the result of one or more errors.

16. The system of claim 11, wherein calculating the likelihood that changes to the set of input data are the result of one or more errors comprises:

17. The system of claim 16, wherein calculating the information content of the input data is performed by calculating the Shannon entropy of the input data.

19. The system of claim 16, wherein the statistical analysis is performed using Bayesian statistics.

20. The system of claim 16, wherein the statistical analysis is performed using parametric statistics.

21. A system for detecting abnormalities in input data to a financial risk management system, the system comprising:

a means for receiving a set of input data to a financial risk management system;

a means for receiving one or more historical values, each historical value representing a calculated content from a previous set of input data; and

a means for calculating the likelihood that changes to the set of input data are the results of one or more errors.

22. The system of claim 21, further comprising:

a graphical user interface means for displaying a result based on the calculated likelihood that changes to the set of input data are the result of one or more errors.

23. A method for detecting abnormalities in data related to a financial risk management system, the method comprising:

(a) receiving a set of data;

(b) receiving one or more historical values, each historical value representing a previous set of data;

(c) calculating the likelihood that changes to the set of data are the result of one or more errors.

24. The method of claim 23, wherein the set of data includes input data to a financial risk management system.

25. The method of claim 23, wherein the set of data includes data calculated by a financial risk management system.

26. The method of claim 23, wherein each value of the one or more historical values represents the information content of a previous set of data.

27. The method of claim 23, wherein calculating the likelihood that changes to the set of data are the result of one or more errors comprises:

- (i) calculating the information content of the data; and
- (ii) performing a statistical analysis of the calculated information content relative to the one or more historical values to determine the likelihood that changes to the data are the result of one or more errors.

24

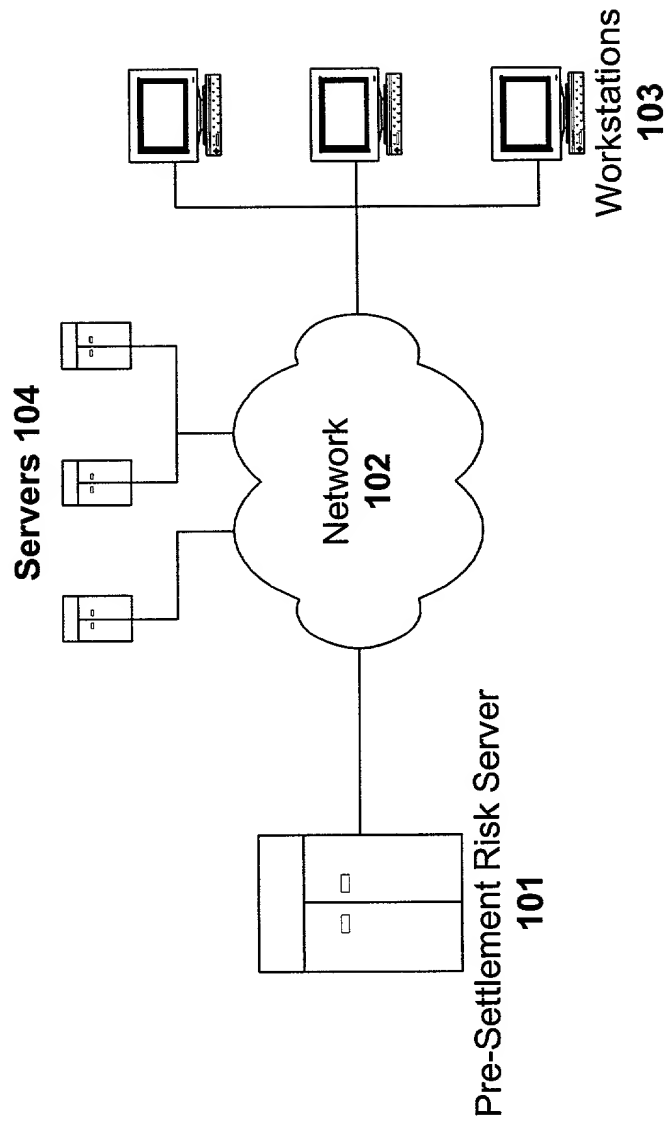


FIG. 1

$\Omega_d(X, Y)$

```

1   $N_{min} \leftarrow 2$ 
2  if  $size[X] < N_{min}$  or  $size[Y] < N_{min}$ 
3      then return (0,0)
4  for each key  $k \in X$ 
5      do  $D_x[X[k]] \leftarrow D_x[X[k]] + 1$ 
6   $MaxEntropy \leftarrow size[X] * \log(size[D_x])$ 
7  for each key  $k \in Y$ 
8      do  $D_y[Y[k]] \leftarrow D_y[Y[k]] + 1$ 
9   $Entropy \leftarrow 0$ 
10 for each key  $k$  of  $D_y$ 
11     do  $Entropy \leftarrow Entropy - D_y[k] * \log(D_y[k]/size[Y])$ 
12 return( $MaxEntropy, Entropy$ )

```

FIG. 2

```

 $\Omega_c(X, Y)$ 
1   $N_{min} \leftarrow 4$ 
2  if  $size[X] < N_{min}$ 
3      then return (0,0)
4   $n_x \leftarrow \lceil \sqrt{size[X]} \rceil$ 
5   $X_{sort} \leftarrow sort[X]$ 
6  for  $j \leftarrow 1$  to  $n_x$ 
7      do  $Bounds[j - 1] \leftarrow X_{sort}[n_x * j - 1]$ 
8   $MaxEntropy \leftarrow n_x * \log(\sqrt{size[X]})$ 
9   $MaxBounds \leftarrow Bounds[n_x - 1]$ 
10 for each key  $v \in Y$ 
11     do for each key  $k$  of  $Bounds$ 
12         do if  $Y[v] < Bounds[k]$ 
13             then  $D_Y[k] \leftarrow D_Y[k] + 1$ ; next
14         if  $Y[v] > MaxBounds$ 
15             then  $D_Y[n_x - 1] \leftarrow D_Y[n_x - 1] + 1$ 
16         next
17 for each key  $k$  of  $D_Y$ 
18     do  $Entropy \leftarrow Entropy - D_Y[k] * \log(D_Y[k]/size[Y])$ 
19 return( $MaxEntropy, Entropy$ )

```

FIG. 3


```

 $\Omega_{\text{cxd}}(X, Y)$ 
1   $N_{\min} \leftarrow 2$ 
2  if  $\text{size}[X] < N_{\min}$  or  $\text{size}[Y] < N_{\min}$ 
3      then return (0,0)
4   $n_x \leftarrow \text{size}[X]$ 
5   $X0_{\text{sort}} \leftarrow \text{sort } X \text{ by 0 dimension}$ 
6  for each key  $k$  of  $X0_{\text{sort}}$ 
7      do  $XD[X0_{\text{sort}}[k][1]] \leftarrow XD[X0_{\text{sort}}[k][1]] + 1$ 
8   $n_m \leftarrow \text{size}[XD]$ 
9   $\alpha \leftarrow \log(n_m)/\log(n_x)$ 
10  $n_c \leftarrow n_x^\alpha$ 
11 if  $n_c < N_{\min}$  or  $n_m < N_{\min}$ 
12     then return(0,0)
13  $\text{Bounds0} \leftarrow \text{FINDBOUNDARIES}(X0_{\text{sort}}, \text{dimension } 0)$ 
14  $\text{MaxEntropy} \leftarrow n_x * \log(n_x^\alpha * n_m)$ 
15  $D_y \leftarrow \text{BINPACK}(Y, \text{Bounds0})$ 
16 for  $i \leftarrow 0$  to  $n_x$ 
17     do for each key  $j$  of  $D_y[i]$ 
18         do  $\text{Entropy} \leftarrow \text{Entropy} - D_y[i][j] * \log(D_y[i][j]/\text{size}[Y])$ 
19 return( $\text{MaxEntropy}$ ,  $\text{Entropy}$ )

```

FIG. 5

$\Omega_{d \times d}(X, Y)$

- 1 $N_{min} \leftarrow 4$
- 2 **if** $size[X] < N_{min}$ **or** $size[Y] < N_{min}$
- 3 **then** return (0,0)
- 4 **for** each key $k \in X$
- 5 **do** $D_{x1}[X[k][0]] \leftarrow D_{x1}[X[k][0]] + 1$
- 6 $D_{x2}[X[k][1]] \leftarrow D_{x2}[X[k][1]] + 1$
- 7 $n_m \leftarrow size[D_{x1}] * size[D_{x2}]$
- 8 $MaxEntropy \leftarrow size[X] * \log(n_m)$
- 9 **for** each key $k \in Y$
- 10 **do** $D_y["Y[k][0]Y[k][1]"] \leftarrow D_y["Y[k][0]Y[k][1]"] + 1$
- 11 $Entropy \leftarrow 0$
- 12 **for** each key k of D_y
- 13 **do** $Entropy \leftarrow Entropy - D_y[k] * \log(D_y[k]/size[Y])$
- 14 return($MaxEntropy, Entropy$)

FIG. 6





Change	Semaphore	Odds favoring problem	Potential of problem (Maximum credible assessment)
Normal		3 to 1	Little potential of problem
Outer Normal		6 to 1	Substantial potential of problem
Borderline		20 to 1	Strong potential of problem
Abnormal		>20 to 1	Decisive potential of problem

FIG. 7

000000 "55442960

Production Global PSR RELEASE CONTROL - Netscape

File Edit View Go Communicator Help

Back Forward Reload Home Search Netscape Print Security Shop

Bookmarks Netsite: <http://dagama.dcs.citicorp.com/>

GED: SSB Assist SSB Access

CITIGROUP

Transactional Content Analysis

NEW YORK 1234567890

[Content Analysis Tutorial](#)

Name	CMTM by Product	MLIV by Product	CMTM by MLIV	Netting by Product	Passed by Product	Failed by Product	Bad by Product
ABC							

Legend

- Normal
- Upper Normal
- Borderline Abnormal
- Abnormal

Return to previous document in history list

FIG. 8

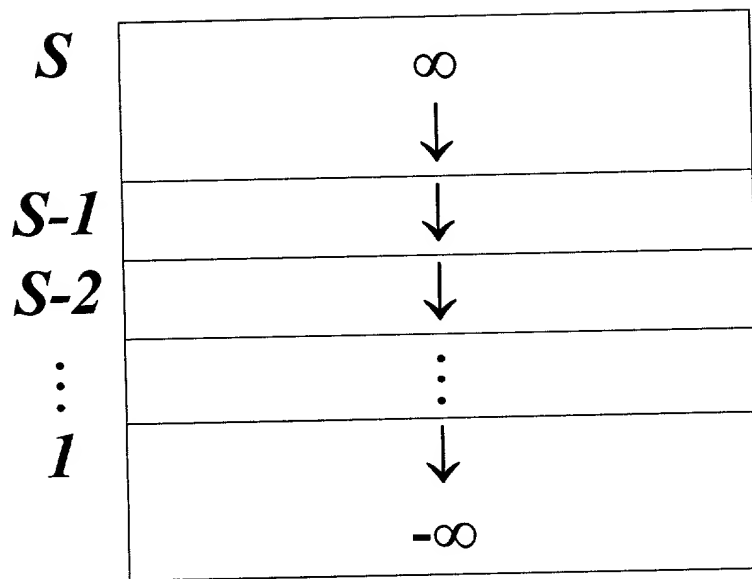


FIG. 9

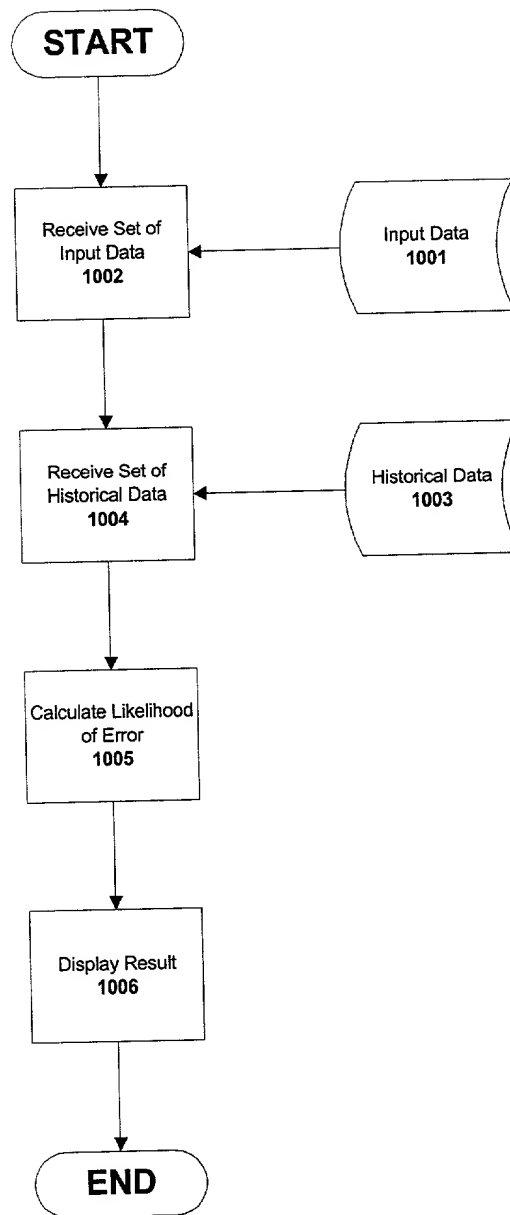


FIG. 10

DECLARATION AND POWER OF ATTORNEY FOR PATENT APPLICATION

English Language Declaration

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name.

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled SYSTEM AND METHOD FOR ASSURING THE INTEGRITY OF DATA USED TO EVALUATE FINANCIAL RISK OR EXPOSURE; the specification of which (check one)

☒ is attached hereto.

☐ was filed on _____ as

☐ Application Serial No. _____

☐ and was amended on _____ (if applicable)

I hereby state that I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose information which is material to patentability as defined in Title 37, Code of Federal Regulations, §1.56.

I hereby claim foreign priority benefits under Title 35, United States Code, §119 of any foreign application(s) for patent of inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate having a filing date before that of the application on which priority is claimed:

Prior Foreign Application(s)

Priority Claimed

(Number)	(Country)	(Day/Month/Year Filed)	Yes	No
----------	-----------	------------------------	-----	----

(Number)	(Country)	(Day/Month/Year Filed)	Yes	No
----------	-----------	------------------------	-----	----

I hereby claim the benefit under Title 35, United States Code, §120 of any United States application(s) listed below and insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35, United States Code §112, I acknowledge the duty to disclose material to patentability as defined in Title 37, Code of Federal Regulations, §1.56 which became available between the filing date of the prior application and the national or PCT international filing date of this application:

60/147,487	August 9, 1999	pending
(Application Serial No.)	(Filing Date)	(Status) (patented, pending, abandoned)

(Application Serial No.)	(Filing Date)	(Status) (patented, pending, abandoned)

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true, and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

English Language Declaration

POWER OF ATTORNEY: As a named inventor, I hereby appoint the following attorney(s) and/or agent(s) to prosecute this application and transact all business in the Patent and Trademark Office connected therewith.

George T. Marcou, Registration No. 33,014; Richard Peterson, Registration No. 35,320; Charles W. Calkins, Registration No. 31,814; John M. Harrington, Registration No. 25,592; A. Jose Cortina, Registration No. 29,733; Russell Todd Morgan, Registration No. 43,815; Charles T. Simmons, Registration No. 35,359; Stephen B. Parker, Registration No. 36,631; James J. Bindseil, Registration No. 42,326; Benjamin Driscoll, Registration No. 41,571; Yoncha L. Kundupoglu, Registration No. 41,130; R. Whitney Winston, Registration No. 44,432; John Ball, Registration No. 44,433; Dawn-Marie Bey, Registration No. 44,442; Tiep Nguyen, Registration No. 44,465, and Greg S. Moldafsky, Registration No. P-46,514.

Send Correspondence to:

George T. Marcou
Kilpatrick Stockton LLP
Suite 800
700 - 13th Street, N.W.
Washington, D.C. 20005

Direct telephone calls to:

George T. Marcou
(202) 508-5800

Full name of sole or first inventor: Ronald Coleman	
First Inventor's Signature	Date
Residence:	14 Scenic Drive, Hyde Park, NY 12538
Citizenship:	USA
Post Office Address:	14 Scenic Drive, Hyde Park, NY 12538
Full name of sole or first inventor: Richard Renzetti	
First Inventor's Signature	Date
Residence:	207 East 27th Street
Citizenship:	USA
Post Office Address:	207 East 27th Street